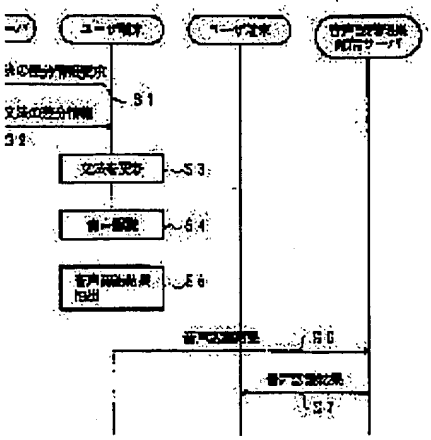


PATENT ABSTRACTS OF JAPAN

(11)Publication number : **2000-172483**
(43)Date of publication of application : **23.06.2000**

(51)Int.CI. **G06F 3/16**
G10L 15/18
G10L 15/28

(21)Application number : **10-351771**
(22)Date of filing : **10.12.1998**

(71)Applicant : **NIPPON TELEGR & TELEPH CORP <NTT>**
(72)Inventor : **USAMI KIYOTADA**
KONO TAKASHI

(54) METHOD AND SYSTEM FOR SPEECH RECOGNITION BY COMMON VIRTUAL PICTURE AND STORAGE MEDIUM STORED WITH SPEECH RECOGNITION PROGRAM BY COMMON VIRTUAL PICTURE

(57)Abstract:
PROBLEM TO BE SOLVED: To enable efficient vocal transmission of an intension to an object in a common virtual picture by allowing a grammar management server to distribute difference information of grammar needed for speech recognition to a user terminal.
SOLUTION: The system has a grammar management server which manages the grammar having a recognition object vocabulary registered in a text, etc., and a speech recognition result distribution server which distributes a speech recognition result. When a user terminal requests the grammar management server to update the difference information of the grammar needed for speech recognition (S1), the grammar management server distributes necessary difference information of the grammar to the user terminal at the start of the speech recognition (S2). The user terminal updates the grammar according to the distributed difference information (S3), recognizes an input voice by using the updated grammar (S4), and extracts (S5) and distributes (S6) the most likelihood vocabulary in the vocabulary in the recognition result to the distribution server. The distribution server

distributes it to other user terminals (S7).

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2000－172483
（P2000－172483A）
(43)公開日　平成12年６月23日(2000.6.23)

(51)Int.Cl.[7]	識別記号	ＦＩ		テーマコード（参考）
Ｇ０６Ｆ　3/16	３２０	Ｇ０６Ｆ　3/16	３２０Ｈ	５Ｄ０１５
Ｇ１０Ｌ　15/18		Ｇ１０Ｌ　3/00	５３７Ｚ	９Ａ００１
15/28			５７１Ａ	

審査請求　未請求　請求項の数８　ＯＬ　（全 10 頁）

(21)出願番号　特願平10－351771

(22)出願日　平成10年12月10日(1998.12.10)

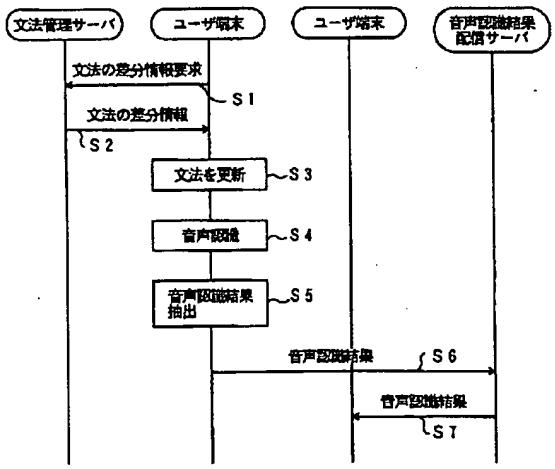
(71)出願人　000004226
日本電信電話株式会社
東京都千代田区大手町二丁目３番１号
(72)発明者　宇佐美　潔忠
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内
(72)発明者　河野　隆志
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内
(74)代理人　100070150
弁理士　伊東　忠彦

Ｆターム(参考)　5D015 GG01 HH23 KK02 LL11
9A001 HH17 HZ32 KK45

(54)【発明の名称】　共有仮想画面における音声認識方法及びシステム及び共有仮想画面における音声認識プログラムを格納した記憶媒体

(57)【要約】
【課題】　ユーザが直接関知していない共有仮想画面内の対象に対して音声による意思伝達を行う際に、効率的に音声による意思伝達を行うために有効な共有仮想画面における音声認識方法及びシステム及び共有仮想画面における音声認識プログラムを格納した記憶媒体を提供する。
【解決手段】　本発明は、ユーザ端末において、文法管理サーバに対して音声認識の際に必要となる文法の差分情報の更新を要求し、文法管理サーバにおいて、ユーザ端末の音声認識処理開始の際に、該音声認識処理に必要な文法の差分情報を該ユーザ端末に配信し、ユーザ端末において、文法管理サーバから配信された差分情報を取得して文法を更新し、更新された文法を用いて、入力された音声の音声認識を行い、音声認識の結果の語彙の中で最尤の語彙を音声認識結果として抽出し、音声認識結果を音声認識結果配信サーバに送信し、音声認識結果配信サーバにおいて、ユーザ端末から送信された音声認識結果を他のユーザ端末に配信する。

本発明の原理を説明するための図



【特許請求の範囲】
【請求項１】　ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能な共有仮想画面における音声認識方法において、
認識対象語彙がテキストなどにより登録された文法を管理する文法管理サーバと、音声認識結果をユーザ端末に配信する音声認識結果配信サーバとを有し、
前記文法管理サーバにおいて、
前記ユーザ端末の音声認識処理開始の際に、該音声認識処理に必要な文法の差分情報を該ユーザ端末に配信することを特徴とする共有仮想画面における音声認識方法。
【請求項２】　前記ユーザ端末において、
音声認識に必要なユーザ独自の音響モデルと言語モデルとを文法と共に格納・管理する請求項１記載の共有仮想画面における音声認識方法。
【請求項３】　前記ユーザ端末において、
入力された音声をユーザ端末において前記文法、前記音響モデル及び前記言語モデルを用いて音声認識する請求項１記載の共有仮想画面における音声認識方法。
【請求項４】　前記ユーザ端末において、音声認識された結果を前記音声認識結果配信サーバに送信する請求項１記載の共有仮想画面における音声認識方法。
【請求項５】　ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能な共有仮想画面における音声認識方法において、
認識対象語彙がテキストなどにより登録された文法を管理する文法管理サーバと、音声認識結果をユーザ端末に配信する音声認識結果配信サーバとを有し、
前記ユーザ端末において、前記文法管理サーバに対して音声認識の際に必要となる文法の差分情報の更新を要求し、
前記文法管理サーバにおいて、前記ユーザ端末の音声認識処理開始の際に、該音声認識処理に必要な文法の差分情報を該ユーザ端末に配信し、
前記ユーザ端末において、前記文法管理サーバから配信された前記差分情報を取得して文法を更新し、
更新された前記文法を用いて、入力された音声の音声認識を行い、
音声認識の結果の語彙の中で最尤の語彙を音声認識結果として抽出し、
前記音声認識結果を音声認識結果配信サーバに送信し、
前記音声認識結果配信サーバにおいて、
前記ユーザ端末から送信された前記音声認識結果を他のユーザ端末に配信することを特徴とする共有仮想画面における音声認識方法。
【請求項６】　ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能な共有仮想画面における音声認識システムであって、
音声認識を行うユーザ端末と、認識対象語彙がテキストなどにより登録された文法を管理する文法管理サーバと、音声認識結果をユーザ端末に配信する音声認識結果配信サーバとを有し、
前記文法管理サーバは、
認識対象語彙がテキストなどにより登録された文法を管理する文法管理手段と、
前記ユーザ端末の音声認識処理開始の際に、該音声認識処理に必要な文法の差分情報を該ユーザ端末に配信する差分情報配信手段とを有し、
前記ユーザ端末は、
音声認識に必要なユーザ独自の音響モデルと言語モデルとを文法と共に格納・管理するモデル格納手段と、
入力された音声をユーザ端末において前記文法、前記音響モデル及び前記言語モデルを用いて音声認識する音声認識手段と、
前記音声認識手段で得られた音声認識結果を送信する音声認識結果送信手段とを有し、
前記音声認識結果配信サーバは、
前記ユーザ端末から送信された音声認識結果を他のユーザ端末に配信する手段を有することを特徴とする共有仮想画面における音声認識システム。
【請求項７】　ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能な文法管理サーバに搭載される共有仮想画面における音声認識プログラムを格納した記憶媒体であって、
前記ユーザ端末の音声認識処理開始の際に、該音声認識処理に必要な文法の差分情報を該ユーザ端末に配信することを特徴とする共有仮想画面における音声認識プログラムを格納した記憶媒体。
【請求項８】　ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能なユーザ端末に搭載される共有仮想画面における音声認識プログラムを格納した記憶媒体であって、
音声認識に必要なユーザ独自の音響モデルと言語モデルとを文法と共に記憶手段に格納・管理するプロセスと、
音声認識を行うために必要な文法を管理する文法管理サーバから配信された文法の差分情報を取得して格納するプロセスと、
入力された音声を前記文法、前記音響モデル及び前記言語モデルを用いて音声認識するプロセスと、
音声認識された結果を前記音声認識結果配信サーバに送信させるプロセスとを有することを特徴とする共有仮想画面における音声認識プログラムを格納した記憶媒体。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、共有仮想画面における音声認識方法及びシステム及び共有仮想画面における音声認識プログラムを格納した記憶媒体に係り、特に、ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能な共有仮想画面において、その共有仮想画面を介して通信される音声を認識するシステム

において、例えば、ユーザが直接関知していない共有仮想画面内の対象に対して音声による意思伝達を行う際に、効率的に音声による意思伝達を行うために有効な共有仮想画面における音声認識方法及びシステム及び共有仮想画面における音声認識プログラムを格納した記憶媒体に関する。

【０００２】

【従来の技術】従来、複数のユーザ端末がネットワークを介してサーバ装置に接続され、複数のユーザが３次元ＣＧによって構築された３次元仮想空間に各々のアバタを参加させることができる３次元共有仮想空間としては、例えば、InterSpace(http://www.ntts.com/ispace.html）がある。このシステムにおいて、ユーザはマイクなどの音声入力デバイスを用いて、３次元共有仮想空間内に参加する他のユーザと音声によるコミュニケーションを行うことが可能である。

【０００３】一方、人間の話す言葉をコンピュータに理解させる音声認識の技術は、人間と機械との情報のやり取りなどを行うためには重要な技術であり、マルチメディアの操作性などのために音声認識技術に対する期待は大きい。ネットワークを介した音声認識システムとしては、クライアントから入力された音声データをサーバに送信し、サーバで音声認識処理を実行した後、その音声認識結果をクライアントに返信するクライアント・サーバ・システムがある。連続音声認識のような処理量の大きなタスクを扱う場合や、即時性を要求するようなタスクを処理する場合に、高速（かつ高価）なワークステーション等をサーバに利用し、低価格なパーソナルコンピュータ上の複数のクライアントから利用するような形態を可能とする。

【０００４】一般的な音声認識処理は、言語音声を構成する要素である音素（母音や子音など）を単位とし、各音素毎に音響的な性質をモデル化した音響モデル、音声認識処理における探索空間を規定するための語彙や文法、または、言語的な単位の統計的な連鎖確率のモデルである言語モデル、及び認識対象語彙がテキストなどにより登録された文法を用いて行われる。入力された音声データ（音声波形）を音声分析し、音声認識に有効な音声特徴量を抽出し、音声認識処理における探索空間を言語モデルにより限定し、音響モデルと音響的な照合を行った後、文法に登録されている語彙の中で尤度の最も高い語彙を音声認識結果として得る。

【０００５】

【発明が解決しようとする課題】最近では、パーソナルコンピュータの処理能力の高まりと音声認識処理アルゴリズムに対する造詣の深まりによる音声認識処理プログラムの簡潔化により、パーソナルコンピュータ上で音声認識処理も行うことができるようになってきている。

【０００６】しかしながら、依然として不特定話者用の音響モデルを利用しており、特定の話者用に作成された音響モデルと同等の音声認識性能を得ることが難しくなっている。また、人間の操作を離れた対象に対して音声で意思伝達をすること（例えば、犬キャラクタに対して「お手」や「お座り」などの音声で命令すること）は困難である。

【０００７】このように、従来のネットワークを介して音声認識技術では、複数のユーザの音声を認識するために、不特定話者用の音響モデルが音声認識サーバには用意されているが、例え、不特定話者用の音響モデルとは言っても、音声の個人差による多様な要因を全て吸収した音声認識を行うことは困難であり、特定の話者用に作成された音響モデルと同等の音声認識性能を得ることは困難である。

【０００８】また、従来の共有仮想画面内における音声によるコミュニケーションは、ネットワークを介して伝播された音声を直接人間が聞き、理解しながらそれに対するリアクションを人間が直接行うという形態を採っている。ここには、人間のリアルタイムな介入が必要不可欠であり、例えば、上記のような人間の操作を離れた犬キャラクタのような対象に対して「お手」や「お座り」などの音声による命令を実現することは困難である。

【０００９】従って、上記の従来の技術には、ユーザが直接関知していない共有仮想画面内の対象に対して音声による意思伝達を行う際に、効率的に音声による意思伝達を行うことができないという問題がある。本発明は、上記の点に鑑みなされたもので、ユーザが直接関知していない共有仮想画面内の対象に対して音声による意思伝達を行う際に、効率的に音声による意思伝達を行うために有効な共有仮想画面における音声認識方法及びシステム及び共有仮想画面における音声認識プログラムを格納した記憶媒体を提供することを目的とする。

【００１０】

【課題を解決するための手段】本発明（請求項１）は、ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能な共有仮想画面における音声認識方法において、認識対象語彙がテキストなどにより登録された文法を管理する文法管理サーバと、音声認識結果をユーザ端末に配信する音声認識結果配信サーバとを有し、文法管理サーバにおいて、ユーザ端末の音声認識処理開始の際に、該音声認識処理に必要な文法の差分情報を該ユーザ端末に配信する。

【００１１】本発明（請求項２）は、ユーザ端末において、音声認識に必要なユーザ独自の音響モデルと言語モデルとを文法と共に格納・管理する。本発明（請求項３）は、ユーザ端末において、入力された音声をユーザ端末において文法、音響モデル及び言語モデルを用いて音声認識する。本発明（請求項４）は、ユーザ端末において、音声認識された結果を音声認識結果配信サーバに送信する。

【００１２】図１は、本発明の原理構成図である。本発

明（請求項５）は、ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能な共有仮想画面における音声認識方法において、認識対象語彙がテキストなどにより登録された文法を管理する文法管理サーバと、音声認識結果をユーザ端末に配信する音声認識結果配信サーバとを有し、ユーザ端末において、文法管理サーバに対して音声認識の際に必要となる文法の差分情報の更新を要求し（ステップ１）、文法管理サーバにおいて、ユーザ端末の音声認識処理開始の際に、該音声認識処理に必要な文法の差分情報を該ユーザ端末に配信し（ステップ２）、ユーザ端末において、文法管理サーバから配信された差分情報を取得して文法を更新し（ステップ３）、更新された文法を用いて、入力された音声の音声認識を行い（ステップ４）、音声認識の結果の語彙の中で最尤の語彙を音声認識結果として抽出し（ステップ５）、音声認識結果を音声認識結果配信サーバに送信し（ステップ６）、音声認識結果配信サーバにおいて、ユーザ端末から送信された音声認識結果を他のユーザ端末に配信する（ステップ７）。

【００１３】図２は、本発明の原理構成図である。本発明（請求項６）は、ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能な共有仮想画面における音声認識システムであって、音声認識を行うユーザ端末１００と、認識対象語彙がテキストなどにより登録された文法を管理する文法管理サーバ２００と、音声認識結果をユーザ端末１００に配信する音声認識結果配信サーバ３００とを有し、文法管理サーバ２００は、認識対象語彙がテキストなどにより登録された文法を管理する文法管理手段１０と、ユーザ端末の音声認識処理開始の際に、該音声認識処理に必要な文法の差分情報を該ユーザ端末に配信する差分情報配信手段２０とを有し、ユーザ端末１００は、音声認識に必要なユーザ独自の音響モデルと言語モデルとを文法と共に格納・管理するモデル格納手段３０と、入力された音声をユーザ端末において文法、音響モデル及び言語モデルを用いて音声認識する音声認識手段４０と、音声認識手段４０で得られた音声認識結果を送信する音声認識結果送信手段５０とを有し、音声認識結果配信サーバ３００は、ユーザ端末１００から送信された音声認識結果を他のユーザ端末に配信する手段を有する。

【００１４】本発明（請求項７）は、ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能な文法管理サーバに搭載される共有仮想画面における音声認識プログラムを格納した記憶媒体であって、ユーザ端末の音声認識処理開始の際に、該音声認識処理に必要な文法の差分情報を該ユーザ端末に配信する。

【００１５】本発明（請求項８）は、ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することが可能なユーザ端末に搭載される共有仮想画面における音声認識プログラムを格納した記憶媒体であって、音声認識に必要なユーザ独自の音響モデルと言語モデルとを文法と共に記憶手段に格納・管理するプロセスと、音声認識を行うために必要な文法を管理する文法管理サーバから配信された文法の差分情報を取得して格納するプロセスと、入力された音声を文法、音響モデル及び言語モデルを用いて音声認識するプロセスと、音声認識された結果を音声認識結果配信サーバに送信させるプロセスとを有する。

【００１６】上記のように、本発明は、認識対象語彙集がテキストなどにより登録された文法を管理する文法管理サーバと、音声認識結果をユーザ端末に配信する音声認識結果配信サーバとを有し、ある音声認識に対する個人的に作成された文法を多数のユーザ端末で共有するために設けられている文法管理サーバにおいて、ユーザ端末の音声認識処理開始の際に、音声認識処理に必要な文法の差分情報をユーザ端末に配信し、得られた差分情報を更新した文法を用いて、入力された音声の音声認識を行い、音声認識された結果を音声認識結果配信サーバに送信し、音声認識配信サーバにおいて、送信された音声認識結果を他のユーザ端末に配信することにより、ユーザが眺めるシーンを描画更新することが可能となる。

【００１７】

【発明の実施の形態】図３は、本発明の音声認識システムの構成を示す。同図に示すシステムは、ユーザ端末１００、文法管理サーバ２００、音声認識結果配信サーバ３００及びネットワーク４００から構成され、ユーザ端末１００、文法管理サーバ２００、及び音声認識結果配信サーバ３００はネットワーク４００に接続されている。

【００１８】ユーザ端末１００は、データ管理部１１０、文法作成処理部１２０、音響モデル適応部１３０、音声認識処理部１４０、文法更新要求部１５０、表示部１６０及び通信制御部１７０から構成される。なお、当該ユーザ端末１００は、図３においては説明の簡単化のため１台がネットワークに接続されているが、本来は、ｎ台が接続されているものとする。

【００１９】データ管理部１１０は、音声認識処理に必要な音響モデル、言語モデル、文法などのデータを格納・管理する。文法作成処理部１２０は、認識対象語彙がテキストなどにより登録されたユーザ独自の文法を作成する。音響モデル適応部１３０は、音声を入力する１８０ユーザの音響特性に合わせて不特定話者用の音響モデルをユーザ用に適応させる。適応させる具体的な方法は、音響モデル適応部１３０により表示部１６０に表示された指示に従い、ユーザ１８０は、不特定話者用の音響モデルにある母音・子音を全て網羅するような発音を行い、音響モデル適応部１３０は、その発音からユーザ１８０の音響特性を捉えて不特定話者用の音響モデルをユーザ用に適応させる。

【００２０】音声認識処理部１４０は、データ管理部１

１０により管理された音声認識処理に必要なデータを用いて、音声認識結果を作成する。文法更新要求部１５０は、音声認識処理部１４０により音声認識を行う対象に対する文法の差分情報の更新を文法管理サーバ２００に対して要求する。表示部１６０は、ユーザ１８０に対して共有仮想シーンを描画更新する。

【００２１】通信制御部１７０は、文法管理サーバ２００や音声認識結果配信サーバ３００との通信を管理する。文法管理サーバ２００は、ある音声認識対象に対するユーザが個人的に作成された文法を多数の他のユーザ端末で共有するため、当該文法を管理するサーバであり、文法データベース２１０、検索部２２０、ロード部２３０、アップロード部２４０及び通信制御部２５０から構成される。

【００２２】文法データベース２１０は、ユーザ端末１００から送信された文法情報を格納・管理する。検索部２２０は、ユーザ端末１００の文法更新要求部１４０から指示された要求に基づいて文法データベース２１０を検索する。ロード部２３０は、検索部２２０で検索されたユーザ端末１００のデータ管理部１１０により管理される文法との差分情報をロードする。

【００２３】アップロード部２４０は、ユーザ端末１００から送信された文法情報を文法データベース２１０にアップロードする。通信制御部２５０は、ユーザ端末１００との通信を管理する。音声認識結果配信サーバ３００は、ユーザ端末１００から取得した音声認識結果を選択した他のユーザ端末に配信するサーバであり、音声認識結果管理部３１０、配信ユーザ管理部３２０、通信制御部３３０から構成される。

【００２４】音声認識結果管理部３１０は、ユーザ端末１００から送信された音声認識結果を格納・管理する。配信ユーザ管理部３２０は、音声認識結果管理部３１０で管理する音声認識結果を配信するユーザ端末１００を管理する。配信するユーザ端末の選択は、特願平１０－３４７３９６に開示されているマスタ端末の選択方法に依るものとする。

【００２５】通信制御部３３０は、ユーザ端末１００との通信を管理する。次に、本発明の動作の概要を説明する。音声認識処理で必要とする音響モデル、言語モデル及び文法の中で共通的な規則により言語モデルと不特定話者用でなく個人の音響特性を反映した音響モデルは個々のユーザ端末１００に蓄積し、他のユーザが共有する個々のユーザが作成した文法は、当該文法を作成したユーザ端末１００自体に蓄積すると共に、他のユーザが共有できるように文法管理サーバ２００に蓄積する。

【００２６】ユーザ端末１００の文法作成処理部１２０により個々のユーザが作成した文法は、ユーザ端末１００の通信制御部１７０、文法管理サーバ２００の通信制御部２５０及びアップロード部２４０を介して文法データベース２１０に更新・蓄積される。なお、共通的な規則の文法は、既にユーザ端末１００のデータ管理部１１０に蓄積されるようにしても良いし、ユーザ端末１００のデータ管理部１１０と文法データベース２１０との両方に蓄積するようにしてもよい。

【００２７】個々のユーザ端末への最新文法の反映処理は、以下のようにして行うものとする。第１の文法の反映処理として、音声認識対象をユーザ１８０が選択し、当該ユーザ端末１００から文法更新要求を発行した場合について説明する。この場合の文法更新は、音声認識対象に関わる文法の差分情報だけを更新する。ユーザ端末１００から音声認識対象に関する文法バージョンを文法管理サーバ２００に送信する。

【００２８】次に、文法管理サーバ２００では、送信された音声認識対象に関する文法バージョンと文法データベース２１０に蓄積されている当該音声認識対象に関する最新の文法バージョンとを比較する。異なっている時だけ、文法データベース２１０の最新バージョンの文法との差分情報をユーザ端末１００に返信する。第２の文法の反映処理として、ワールド全体に関して共通の文法が必要な場合（例えば、関西の文化圏を反映させたワールドでは関西弁が通じる必要がある）に、文法管理サーバ２００からユーザ端末１００に対して文法更新を行う方法について説明する。

【００２９】この場合の文法更新時期は、例えば、ユーザ端末１００がワールド管理サーバ（ワールドへのユーザのログイン、ログアウトを管理するサーバ（図示せず））を介してワールドに参加した際に、全体の文法バージョンを送信し、文法管理サーバ２００では、ワールド管理サーバ（図示せず）経由で送付された全体の文法バージョンを比較し、異なっている時に、最新の文法情報全体をユーザ端末１００に送信する。

【００３０】第３の文法の反映処理として、音声認識対象の方からユーザのアバタに近づいてきて、受動的にそれに必要な文法更新を行う方法について説明する。この場合の文法更新方法は、例えば、音声認識対象が自分に対する文法バージョンを伴っていて、ユーザのアバタにある範囲の距離に近づいた時に、ユーザ端末１００に蓄積されている音声認識対象に対する文法バージョンとを比較し、異なっている場合は、ユーザ端末の文法バージョンを文法管理サーバ２００に送信し、文法差分情報を得て更新する。または、ユーザのアバタにある範囲の距離に近づいた時を契機として、前述の第１の文法の反映処理を行う。

【００３１】次に、上記の構成による音声認識処理の動作を説明する。図４は、本発明の音声認識処理のフローチャートである。

ステップ１０１）　ユーザ端末１００は、文法更新要求部１５０により、音声認識を行う際に必要となる文法の差分情報の更新を要求し、文法の更新を行う必要があるか否かを判定する。更新を行う必要がある場合には、ス

テップ１０２に移行し、必要がない場合にはステップ１０３に移行する。

【００３２】ステップ１０２）　文法管理サーバ２００のロード部２３０によりロードされた文法の差分情報に基づいて、ユーザ端末１００のデータ管理部１１０により管理された文法を更新する。

ステップ１０３）　次に、ユーザ端末１００のデータ管理部１１０により管理された音声認識処理に必要なデータを用いて音声認識処理部１４０により音声認識を行う。

【００３３】ステップ１０４）　そして、ユーザ端末１００のデータ管理部１１０により管理された文法に登録されている語彙の中で尤度の最も高い語彙をテキスト形式の音声認識結果として抽出する。

ステップ１０５）　最後に、テキスト形式にて表現された音声認識結果を音声認識結果配信サーバ３００送信し、本処理を終了する。

【００３４】

【実施例】以下、図面と共に本発明の実施例を説明する。本実施例の前提として、ユーザ端末１００のデータ管理部１１０は、命令コマンドと当該命令に対応するアクションが記載されているアクションテーブル６００を有し、当該ユーザ端末１００がマスタ端末となった場合（マスタ端末となった場合の当該ユーザ端末の処理は、特願平１０－３４７３９６号に詳述されている）には、アクションテーブル６００に記載されているアクションに対応する処理を行うものとする。

【００３５】以下の実施例では、共有仮想画面として、３次元共有仮想空間を例にとって説明する。図5は、本発明の一実施例の３次元共有仮想空間における音声認識の例を示す。同図に示すように、実際に、ユーザ端末１００を用いてあるワールド５００に参加するユーザ５１０が、当該ワールド５００内に存在する犬キャラクタ５２０に対して音声による意思伝達を行う場合について説明する。なお、この犬キャラクタ５２０は、図6に示すようなアクションテーブル６００に基づいて動作を行うものとする。

【００３６】図7は、本発明の一実施例の文法データベースの構成を示す。同図に示す文法データベース２１０は、音声認識対象毎に１つのファイルが割り当てられた複数のファイルで構成される。同図では、１つのファイルに１つの文法が格納されている状態を表しており、文法識別子と開始は１つの文法の区切りであり、多数の文法の一つ一つが文法識別子と開始に囲まれて１つのファイルに格納されている。

【００３７】単語宣言は、１つの文法に使われる有意義な単語を全て定義しており、音響モデル及び言語モデルを利用して認識された単語が有意義な単語にどのように対応するか、また、同じ意味の単語はあるかを表している。この例では、同じ意味の単語は、「｜」印のＯＲ記号により表されており、「いどう」と「いどうしてください。」は同じ有意義な単語としている。

【００３８】文法宣言は、意味のある１つの文として、どのように有意義な単語群からどのような順序配列で構成されているかを表している。この文法の表記方法は既に知られている方法である。この例では、

・$number=($1 | $2 | $3);「１」または、「２」または、「３」は『ｎｕｍｂｅｒ』とする。

・$num=$number$ 歩；「１歩」または、「２歩」または、「３歩」は『ｎｕｍ』とする。

・$direction=$前；「前」は『ｄｉｒｅｃｔｉｏｎ』とする。

・$dir=$direction$に；「前に」は『ｄｉｒ』とする。

・number=(numdir | dirnum) ；「１歩」または、「２歩」または、「３歩」かつ「前に」または、「前に」から「１歩」または「２歩」または、「３歩」は『ｎｕｍｂｅｒ』とする。

・$sentence=$numdir$移動；「１歩」または、「２歩」または、「３歩」かつ「前に」または、「前に」から「１歩」または「２歩」または、「３歩」かつ「移動」は『ｓｅｎｔｅｎｃｅ』とする。

【００３９】従って「１歩前に移動」も「前に１歩移動」も「１歩前に移動して下さい」も同じ意味となる。また、１つのファイル毎に２つのバージョン情報があり、また、文法識別子と開始とで区切られた文法毎にバージョン情報がある。これらを利用する場合には、文法データベース２１０の１ファイル毎にある２つのバージョン情報として、一つは、１ファイル全体を文法管理サーバ２００を維持・管理する制御装置等からの制御により更新するときに、更新・利用されるバージョン情報である。このバージョン情報は、バージョン番号と作成日等からなり、ワールド全体に関して共通の文法が必要な場合（関西の文化圏を反映させたワールドでは関西弁が通じる必要がある）に、文法管理サーバ２００からユーザ端末１００に対して文法全体の更新を行うときに利用される。

【００４０】もう一つの文法データベース２１０の１ファイル毎にあるバージョン情報は、最新の更新日・時刻等からなり、ユーザ端末１００からの文法作成による文法データベース２１０のファイル更新毎に、その更新日・時刻に更新される。このバージョン情報は、ユーザ端末１００の文法更新要求部１５０からデータベースと構成フォーマットが同じであるデータ管理部１１０に記録されている音声認識対象に対してのバージョン情報を送信する時に利用されるものである。このバージョン情報は、個々の文法にあるバージョン情報から検索処理により、最新の更新日・時刻を得る処理よりも高速に最新の更新日・時刻を取得するために用いるものであるが、ユーザ端末１００の処理能力が高く、当該バージョン情報を用いずに文法毎にあるバージョン情報の検索処理をし

ても容易に最新の更新日・時刻を取り出すことができて、他の処理に影響を与えないとするならばこのファイル毎のバージョン情報は無くてもよい。

【００４１】文法毎にあるバージョン情報は、ユーザ端末１００の文法作成処理部１２０で音声認識対象に対して作成された文法が送信される毎に更新される。このバージョン情報は、最新の更新日・時刻（送信された日）等からなっている。また、作成文法を送信したユーザ端末１００のデータ管理部１１０においても、作成文法と共に、バージョン情報が送信時に更新される。この時のユーザ端末１００における文法作成時のデータ管理部１１０の作成文法及びバージョン情報の更新方法は、単に作成した文法及びそのバージョン情報だけのユーザ端末１００の処理による更新では、他の文法のバージョンが旧世代のままになってしまう恐れもある（なぜなら、音声認識対象に対する全体文法の最新の更新日・時刻であるバージョン情報は、最新になるのに、他の文法は、依然と旧世代のままの内容であり、修正前の最新更新日の日付から修正日の最新更新日までの変化が反映されなくなる）ので、データ管理部１１０に記憶されている修正前の音声認識対象に対する全体文法の最新の更新日・時刻であるバージョン情報も送信し、下記に示す文法管理サーバ２００の処理により差分情報を得て更新する必要がある。

【００４２】なお、文法管理サーバ２００は、文法データベース２１０の文法毎にあるこのバージョン情報を、ユーザ端末１００の文法更新要求部１５０から送信されてきたデータ管理部１１０に記憶されている音声認識対象に対してのバージョン情報（最新の更新日・時刻）と比較して、より新しい更新日・時刻のバージョンの文法データベース２１０にある文法だけを差分情報として、ユーザ端末１００に送信する。ユーザ端末１００では、送信された音声認識対象に対する文法の差分情報でデータ管理部１１０の個々の文法を更新し、最新のものとする。同じく、バージョン情報（最新の更新日・時刻）も更新する。

【００４３】図８は、本発明の一実施例の音声認識処理の動作の例を示すシーケンスチャートである。

ステップ２０１）　図５において、ワールド５００内の犬キャラクタ５２０をマウスによりクリックすることなどにより音声認識処理を開始すると、ユーザ端末１００の文法更新要求部１５０により、文法管理サーバ２００に対して、犬キャラクタ５２０に対する音声認識処理を行う際に必要となる文法の差分情報の更新を要求する。

【００４４】ステップ２０２）　更新要求を受信した文法管理サーバ２００は、検索部２２０により、その犬キャラクタ５２０に対する音声認識処理を行うために必要な文法を文法データベース２１０から検索し、ロード部２３０により、ユーザ端末１００のデータ管理部１１０により管理される文法との差分情報をロードする。

ステップ２０３）　次に、ロード部２３０によりロードされた文法の差分情報は、ユーザ端末１００のデータ管理部１１０により更新・格納され、この文法を用いて実際の音声認識処理が可能となる。なお、ここでは、音声認識対象をユーザが選択した際のクライアント側（ユーザ）からの文法更新要求を例としているが、例えば、関西の文化圏を反映させたワールドでは関西弁が通じるように、あるワールド全体に関して共通の文法が必要な場合には、サーバ側（文法管理サーバ２００）からクライアント（ユーザ）に対して文法更新を行ったり、音声認識対象の方から近づいてきて受動的にそれに必要な文法更新を行うなど、色々な適用例が考えられる。

【００４５】ここで言う文法とは、ユーザが発声する音声を仮名で記述した文字列とそれに対応する文字列の表記との組み合わせのことを指す。例えば、図６に示すようなアクションテーブル６００において、「１歩前に移動」というテキストコマンドを音声認識結果として得る場合には、文法の中には最低限「いっぽまえにいどう」という発声が「１歩前に移動」という文字列の表記に対応する、と登録されていなければ認識することはできない。また、「１歩前に移動」も「前に１歩移動」も「１歩前に移動して下さい」という発声も音声認識結果としては「１歩前に移動」と同一である、と文法の中に登録されていなければ、それぞれが異なる音声認識結果を得ることになってしまう。これらの発声は内容的には皆同じではあるが、文法の登録の仕方によっては異なる音声認識結果を得ることに繋がる。逆に全く同じ発声ではあっても、文法の登録のしかたによって、異なる音声認識結果を得ることができるので、音声認識処理を行う対象毎に異なった文法を登録して、その文法をユーザ端末に逐次反映することにより、ユーザの音声による意思伝達を柔軟に行うことが可能となる。

【００４６】ステップ２０４）　犬キャラクタ５２０に対する音声認識処理は、ユーザ端末１００の文法更新要求部１４０の更新要求により更新された文法を用いて、音声認識処理部１３０により行われる。

ステップ２０５）　テキスト形式にて表現された音声認識結果は、音声認識結果配信サーバ３００に送信される。

【００４７】ステップ２０６）　音声認識結果を受信した音声認識結果配信サーバ３００は、音声認識結果管理部３１０により、受信した音声認識結果を一旦格納する。

ステップ２０７）　次に、配信ユーザ管理部３２０により、受信した音声認識結果を配信するユーザ端末を選択する。

ステップ２０８）　通信制御部３３０により、選択されたユーザ端末１００に対して音声認識結果の配信を行う。

【００４８】ステップ２０９）　音声認識結果配信サー

バ３００から配信された音声認識結果を受信したユーザ端末１００は、アクションテーブル６００により定義された該当するアクションを犬キャラクタ５２０に対して開始させる。これらにより、ユーザが直接関知していない共有仮想空間内の対象に対して音声による意思伝達を行う際に、効率的に音声による意思伝達を行うことが可能となる。

【００４９】ステップ２１０）　一方、ユーザ端末１００の文法作成部１２０により作成されたユーザ独自の文法は、データ管理部１１０により更新・格納される。

ステップ２１１）　同時に、通信制御部１７０により、文法管理サーバ２００に送信される。

ステップ２１２）　文法情報を受信した文法管理サーバ２００は、アップロード部２４０により、文法データベース２１０にその文法情報をアップロードする。

【００５０】これにより、ユーザ独自の文法は常に最新の文法として他のユーザ端末に反映させることが可能となる。ここで、上記のステップ２０９におけるアクションテーブル６００の定義に基づいてアクションを起こす場合について説明する。犬キャラクタ５２０に対する命令を認識したユーザ端末１００は、その音声認識した命令を音声認識結果配信サーバ３００に送信する。音声認識結果配信サーバ３００は、その犬キャラクタ５２０に対する命令を犬キャラクタ５２０の行動を管理するマスタ端末に配信する。犬キャラクタの行動を管理するマスタ端末では、マスタ端末の共有オブジェクト処理手段（図示せず）がデータ管理部に記憶されているアクションテーブルを利用して、犬キャラクタに対する命令に対応する行動を行わせる。

【００５１】また、音声認識処理に必要となる音響モデルをユーザ個別に管理することにより、実際に音声を入力するユーザ毎に音響モデルを柔軟に適応させることができるので、ユーザ毎に最高の音声認識性能を得ることが可能となる。このように、本発明によれば、ユーザの視点に応じて、ユーザが眺めるシーンを描画更新することができる共有仮想画面において、ユーザが直接関知していない共有仮想画面内の対象に対して音声による意思伝達を行う際に、効率的に音声による意思伝達を行うことができる。

【００５２】また、上記の実施例は、図３の構成に基づいて説明しているが、ユーザ端末１００、文法管理サーバ２００及び音声認識結果配信サーバ３００の各構成要素をプログラムとして構築し、それぞれの端末、サーバとして利用されるコンピュータに接続されるディスク装置や、フロッピーディスク、ＣＤ－ＲＯＭ等の可搬記憶媒体に格納しておき、本発明を実施する際にインストールすることにより、容易に本発明を実現できる。

【００５３】なお、本発明は、上記の実施例に限定されることなく、特許請求の範囲内で種々変更・応用が可能である。

【００５４】

【発明の効果】上述のように、本発明では、共有仮想画面における音声認識において、ユーザの視点に応じてユーザが眺めるシーンを描画更新することが可能な共有仮想画面において、ユーザが直接関知していない共有仮想画面内の対象に対して音声による意思伝達を行う際に、効率的に音声による意思伝達を行うことができる。

【図面の簡単な説明】

【図１】本発明の原理を説明するための図である。

【図２】本発明の原理構成図である。

【図３】本発明の音声認識システムの構成図である。

【図４】本発明の音声認識処理のフローチャートである。

【図５】本発明の一実施例の３次元共有仮想空間における音声認識の例である。

【図６】本発明の一実施例のアクションテーブルの例である。

【図７】本発明の一実施例である文法の例である。

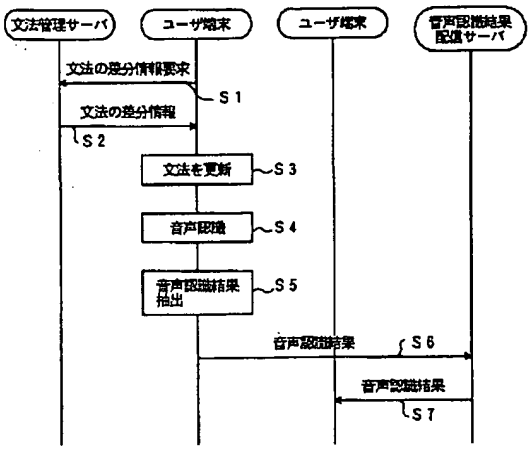
【図８】本発明の一実施例の音声認識の動作の例を示すシーケンスチャートである。

【符号の説明】

１０　文法管理手段
２０　差分情報配信手段
３０　モデル格納手段
４０　音声認識手段
５０　音声認識結果送信手段
１００　ユーザ端末
１１０　データ管理部
１２０　文法作成処理部
１３０　音響モデル適応部
１４０　音声認識処理部
１５０　文法更新要求部
１６０　表示部
１７０　通信制御部
１８０　ユーザ
２００　文法管理サーバ
２１０　文法データベース
２２０　検索部
２３０　ロード部
２４０　アップロード部
２５０　通信制御部
３００　音声認識結果配信サーバ
３１０　音声認識結果管理部
３２０　配信ユーザ管理部
３３０　通信制御部

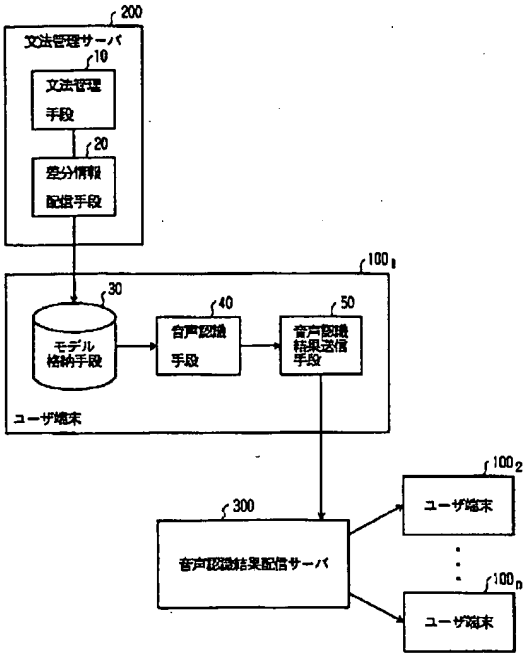
【図1】

本発明の原理を説明するための図



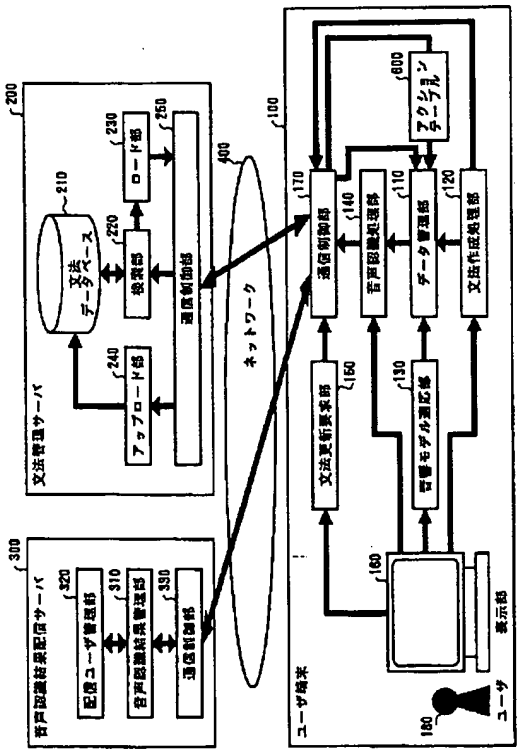
【図2】

本発明の原理構成図



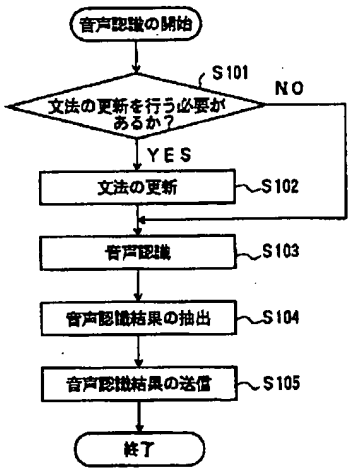
【図3】

本発明の音声認識システムの構成図



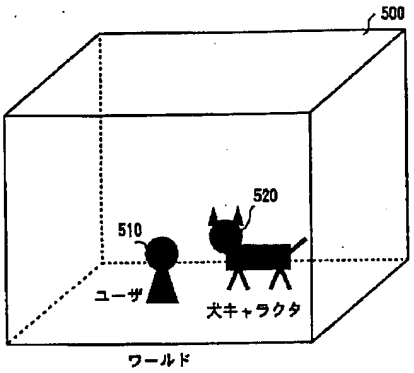
【図4】

本発明の音声認識処理のフローチャート



【図5】

本発明の一実施例の3次元共有仮想空間における音声認識の例



【図6】

本発明の一実施例のアクションテーブルの例

800

テキストコマンド	アクション
「1歩前に移動」	1歩前に移動する動作
「90度左に回転」	90度左に回転する動作
・	
・	
・	

【図7】

本発明の一実施例の文法データベースの構成図

```
/* 文法識別子 */
BNF

/* 単語宣言 */
$[1] = (いち|いっ);
$[2] =に;
$[3] =さん;
$[歩] = (ほ|ぽ);
$[前] =まえ;
$[に] =に;
$[移動] = (いどう|いどうしてください);

/* 文法宣言 */
$number=($1|$2|$3);
$num=$number$歩;
$direction=$前;
$dir=$direction$に;
$numdir=($num$dir|$dir$num);
$sentence=$numdir$移動;

/* 開始 */
$START=pause$sentence pause;
```

【図8】

本発明の一実施例の音声認識の動作の例を示すシーケンスチャート